TD n^{o}6

Corrigé

1. b) Une durée de temps t_0 (exprimée en secondes) est représentée par $t_0 * \texttt{freq}$ points. On arrondit ce chiffre à l'entier le plus proche.
e) Pour simplifier, on fait comme si f était un signal infini. Alors la transformée de Fourier de $g[k] = f[k+1] - f[k]$ est :

$$\begin{aligned}
\hat{g}(\omega) &= \sum_{k\in\mathbb{Z}} g[k]e^{-ik\omega} \\
&= \sum_{k\in\mathbb{Z}} (f[k+1] - f[k])e^{-ik\omega} \\
&= \sum_{k\in\mathbb{Z}} f[k]e^{-i(k-1)\omega} - \sum_{k\in\mathbb{Z}} f[k]e^{-ik\omega} \\
&= (e^{i\omega} - 1)\hat{f}(\omega)
\end{aligned}$$

Cette transformation a pour effet d'atténuer un peu les basses fréquences (qui correspondent aux valeurs de ω pour lesquelles $e^{i\omega}$ est proche de 1) par rapport aux plus hautes fréquences. Cela permet de mieux observer les hautes fréquences sur le spectrogramme ; sans cette transformation, elles ont tendance à être d'intensité négligeable par rapport aux basses fréquences.
f) La k-ième fenêtre aura pour support $\{(k-1)\texttt{step\_size}, ..., (k-1)\texttt{step\_size}+\texttt{win\_size}-1\}$. Pour que cet ensemble reste inclus dans $\{0, ..., N-1\}$ (où N est la taille du signal $\texttt{g}$), il faut :

$$k \leq \frac{N - \texttt{win\_size}}{\texttt{step\_size}} + 1$$

i) Soient h est un signal de taille N et H sa transformée de Fourier avec un facteur de sur-échantillonage égal à a. Pour $n = 0, ..., N-1$, la valeur $H[an] = \hat{h}[n]$ correspond à la fréquence pure $k \to \exp\left(2\pi i \frac{kn}{N}\right)$. La période de cette fonction est N/n.
Ici, les signaux dont on a calculé la transformée de Fourier sont de taille $\texttt{win\_size}$ et le facteur de sur-échantillonage est égal à $\texttt{ov}$. La $(\texttt{ov}.k+1)$-ème ligne de la matrice représente une fréquence pure dont la période est de $\texttt{win\_size}/k$ points, ce qui représente $\frac{\texttt{win\_size}}{k\times\texttt{freq}}$ secondes. La fréquence étant l'inverse de la période, cela correspond à la fréquence $\frac{k\times\texttt{freq}}{\texttt{win\_size}}$.
La $k+1$-ème ligne correspond donc à la fréquence suivante, exprimée en Herz :

$$\frac{k \times \texttt{freq}}{\texttt{win\_size} \times \texttt{ov}}$$

j) Il ne faut garder la ligne k que si :

$$\frac{(k-1) \times \texttt{freq}}{\texttt{win\_size} \times \texttt{ov}} \leq 5000$$

c'est-à-dire :

$$k \leq \frac{5000 \times \texttt{win\_size} \times \texttt{ov}}{\texttt{freq}} + 1$$

l) Le signal `gemshorn.wav` représente un morceau de flûte. Chaque note correspond à un ensemble de lignes horizontales sur le spectrogramme. Les hauteurs de ces lignes horizontales correspondent aux différents multiples de la fréquence fondamentale de la note et leur longueur est proportionnelle à la durée de la note.
Le signal `coucou.wav` représente un cri d'oiseau dont la fréquence varie en fonction du temps.
Le signal `chirp.wav` contient deux « chirps » semblables, dont la fréquence commence à 440 Hz puis tend très rapidement vers l'infini. On voit bien cette évolution de la fréquence sur le spectrogramme.
2. b) D'après la construction de la question 1., pour tous k et n, $\texttt{S}[k.\texttt{ov}+1,n]$ est le k-ième point de la transformée de Fourier discrète du signal suivant :

$$\forall m = 0, ..., \texttt{win\_size}-1, \quad s[m] = \texttt{window}[m] \times \texttt{f}[(n-1)\texttt{step\_size}+m]$$

Donc $\texttt{S}[k.\texttt{ov}+1,n]$ vaut :

$$\begin{aligned}
&\sum_{m=0}^{\texttt{win_size}-1} \texttt{window}[m] \times \texttt{f}[(n-1)\texttt{step_size}+m] \times \exp\left(-2\pi i \frac{mk}{\texttt{win_size}}\right) \\
=&a \sum_{m=0}^{\texttt{win_size}-1} \texttt{window}[m] \times \exp\left(2\pi i\nu \frac{(n-1)\texttt{step_size}+m}{\texttt{freq}} - 2\pi i \frac{mk}{\texttt{win_size}}\right) \\
=&a \exp\left(2\pi i\nu \frac{(n-1)\texttt{step_size}}{\texttt{freq}}\right) \sum_{m=0}^{\texttt{win_size}-1} \texttt{window}[m] \times \exp\left(2\pi i \frac{m}{\texttt{win_size}}\left(\frac{\nu.\texttt{win_size}}{\texttt{freq}} - k\right)\right) \\
=&a \exp\left(2\pi i\nu \frac{(n-1)\texttt{step_size}}{\texttt{freq}}\right) \widehat{\texttt{window}}\left[k - \frac{\nu.\texttt{win_size}}{\texttt{freq}}\right]
\end{aligned}$$

c'est-à-dire que $\texttt{S}[k,n]$ vaut :

$$a \exp\left(2\pi i\nu \frac{(n-1)\texttt{step\_size}}{\texttt{freq}}\right) \widehat{\texttt{window}}\left[\frac{k-1}{\texttt{ov}} - \frac{\nu.\texttt{win\_size}}{\texttt{freq}}\right]$$

Puisque $|\widehat{\texttt{window}}|$ est maximale en 0, $|\texttt{S[k,n]}|$ est maximal quand $\frac{k-1}{\texttt{ov}} - \frac{\nu.\texttt{win\_size}}{\texttt{freq}} = 0$, c'est-à-dire :

$$k = \frac{\texttt{ov}.\nu.\texttt{win\_size}}{\texttt{freq}} + 1$$

e) Les fréquences instantanées sont assez bien détectées mais il y a un peu de bruit. En particulier, autour des lignes horizontales les plus basses, on a tendance à détecter des fréquences instantanées parasites, qui dessinent d'autres lignes horizontales. Ce phénomène est dû au faut que la transformée de Fourier de la fenêtre a des maxima locaux ailleurs qu'en 0.
f) D'après la question b), $[k,n]$ correspond à une fréquence instantanée ν si :

$$k = \frac{\nu.\texttt{win\_size}.\texttt{ov}}{\texttt{freq}} + 1$$

Dans ce cas, on a, toujours par la question b) :

$$\frac{\texttt{S}[k, n+1]}{\texttt{S}[k, n]} = \exp\left(2\pi i \nu \frac{\texttt{step\_size}}{\texttt{freq}}\right) = \exp\left(2\pi i (k-1) \frac{\texttt{step\_size}}{\texttt{win\_size.ov}}\right)$$

h) Tant que la fréquence instantanée varie doucement, le résultat est correct. En revanche, lorsque la fréquence instantanée tend vers $+\infty$, l'hypothèse selon laquelle la fréquence instantanée est à peu près constante sur chaque fenêtre n'est plus vraie. La fréquence instantanée n'est donc plus bien détectée.